

# 面向数据共享目的的 GIS 语义数据模型

陈常松

何建邦

(中国测绘科学研究院, 北京 100039) (中科院地理所信息室, 北京 100101)

**摘要** 提出了一种 GIS 语义数据模型, 试图通过该模型从技术上解决(至少是部分地解决)部门之间的数据共享问题。给出了该模型的理论框架, 及有关概念严格的、形式化的定义, 并给出了一个应用实例。最后, 指出了需要继续进行的工作。

**关键词** GIS 语义数据模型 地理实体 地理特征 概念模式

机构之间 GIS 数据的共享涉及许多因素, 例如政策管理、标准等, 但语义层次上的数据共享需要一种语义层次上的数据模型作为共享数据的基础, 一般将这种模型称为全局性概念模型或 GIS 语义数据模型<sup>[1]</sup>。当前相关 GIS 技术及产品中所通常采用的矢量数据模型和栅格数据模型面向纯粹的空间数据(并非仅限于地理数据)管理, 所注重的是空间数据如何表达的问题。面向 GIS 数据共享的语义数据模型强调对地理事实的完整、正确及规范的表达, 其不但关注空间属性的数据表达, 而且更侧重于地理属性数据及其间关系的表达, 其模型方法一般采用面向实体的方法。在面向数据共享的语义数据模型设计中, 诸如“矢量”“栅格”模型仅仅作为这种语义数据模型的一个子模型。本文首先分析了我国传统的用于 GIS 数据共享的手段——主要是地学编码的优缺点, 在此基础上提出一个规范化的 GIS 语义数据模型, 第三部分结合我国正在进行的 NREDIS 的设计给出了一个具体的应用实例。最后给出了需要继续进行的工作要点。

## 1 地理信息分类和编码用于数据共享

地理信息的分类学研究以及建立在其上的信息编码, 是传统地理分析技术在 GIS 中的继续深化, 最

初的目的是便于地理数据(尤其是资源与环境数据)的处理和共享, 但是实践表明这些目的并没有全部达到, 尤其在共享方面。原因主要有下列几个方面:

(1) 建立分类体系时的分类指标不一, 造成了对地理信息编码表示的混乱

一般地讲, 对地理信息分类所采用的指标至少有 3 类<sup>[2]</sup>, 分类指标的不同导致分类结构的不同从而使信息的编码也就不同。同一个地理实体基于不同分类体系可能具有完全不同的编码, 从而影响数据的有效共享; 从应用角度看, 不同的资源与环境管理部门由于采用了不同的分类指标, 建立了适合于自己应用需要的分类体系, 这种分类体系不但会带来不同的信息编码, 而且在数据组织方面导致下列问题: 第一, 同义不同名, 同一地理实体采用完全不同命名; 第二, 同名不同义, 同一名称表达完全不同的地理实体; 第三, 同一地理实体在不同分类体系中处于不同的分类位置。这 3 种状况都影响机构之间的数据共享行为。

(2) 我国已有的对地理要素的编码方案标准, 仍然深受地图表示的影响, 注重于数据的空间表达, 破坏了地理客体的完整性。

例如 GB/T 13923-92<sup>[3]</sup>对河流体系的编码中有 3 个连续的码, 它们是“21011”、“21012”、“21013”, 所代表的含义分别为“单线河”、“双线河左岸”和“双

线河右岸”。毫无疑问,“双线河左岸”和“双线河右岸”并不是“水系”特征中的独立实体,不能代表“河流”的最低划分单位,从分类的角度,这种划分是不合理的。在实际的数据库操作中容易导致错误的产生,例如以这3个码作为关键字进行查询所获得的可能是2条河流而不是3条河流。

(3) 分类体系所采用的是一种层次结构,表达的是不同地理特征之间,不同的地理实体之间的分类隶属关系,不能表达复杂地理空间的全部语义。传统的线分类体系只支持1:n关系,但是地理环境中的分级关系中,上一级类别与下一级类别之间的关系往往是一种n:n关系,部门之间的这种n:n关系对部门之间的数据共享具有特别的意义。线性的分类体系,由于不支持n:n关系,从而也就妨碍了部门之间数据的共享。

## 2 面向数据共享的GIS语义数据模型

### 2.1 空间体系:一种空间概念

分类分级方法应用于地理系统的空间特性研究时,就形成所谓的空间体系或称空间有序性。在所有的空间有序性研究中,地理空间的特化和泛化及地理空间的离散性和连续性研究对GIS数据模型的研究具有重要的意义。

**定义** 直接或间接与地表分布有关的事物、现象、特征等的总和,连同对它们的描述称为地理现实,以点集表示二维空间,则地理现实表现为一多元组<sup>[4]</sup>。

$$\langle x, y, a_1, a_2, \dots, a_n \rangle \quad (1)$$

其中 $x, y$ 表示二维空间的任意点, $\{a_1, a_2, \dots, a_n\}$ 表示在该点上的地理事物、地理特征、地理现象的集合。

**定义** 如果集合 $A = \{a_1, \dots, a_i, \dots, a_n\}$ 表示地理事物 $a_1, a_2, \dots, a_i, \dots, a_n$ 在 $P_i$ 上的综合函数效应, $P_i = (x_i, y_i)$ 表示二维平面上的任意点, $S = \{P_1, \dots, P_i, \dots, P_n\}$ 表示作为研究对象的二维平面,则元组

$$\langle S, A \rangle \quad (2)$$

表示一种泛化的地理空间。由式(2)可知,地理空间是一连续空间,这一空间的地理意义由集合 $A$ 表示。由于地理环境的复杂性,往往难于获取对 $A$ 的精确描述,并且在实践中 $A$ 也不具备太多的意义,因此一般地作为研究对象的地理空间是特化的地理空间。

**定义** 如果存在集合 $U \subset A$ ,则元组

$$\langle S, U \rangle \quad (3)$$

称为关于属性集合 $U$ 的特化空间。特别的,如果 $U = \{a_i\}$ ,则元组 $\langle S, U \rangle$ 表示关于 $a_i$ 的特化空间,记为 $\langle S, a_i \rangle$ 。

实际上,式(3)表示的是集合 $U$ 中的所有元素的某种函数在空间 $S$ 上的分布状况,因此称为空间变量,记为 $V$ ,并且

$$V = f(U) \quad (4)$$

特别地

$$V_i = f(a_i) \quad (a_i \in U) \quad (5)$$

称为关于地理属性 $a_i$ 的空间变量。不失一般性,下文我们将以 $a_i$ 代替 $V$ ,称 $a_i$ 为空间变量,并研究单个元素 $a_i$ 的空间分布特点,如果空间变量是关于 $a_i$ 的集合 $U$ 的函数分布,可以由 $a_i$ 的研究中简单地推论获得相应的结论。

**定义** 如果有空间变量 $a_i \in A$ ,其在点集空间 $S$ 中的某一函数 $f$ 的值的集合记为 $Z$ ,点 $P_i$ 上的值记为 $Z_{P_i} \in Z$ ,即

$$Z_{P_i} = f(a_i, P_i) \quad (a_i \in A, P_i \in S) \quad (6)$$

如果存在任一点集 $P \subset S$ 符合某一规律,并且 $Z_{P_i}$ 在 $P$ 上表现某种离散特征,则称 $S$ 空间为离散空间,如果 $Z_{P_i}$ 在 $P$ 上的值是连续的,则叫做 $S$ 为连续空间,其中 $P$ 要符合的规律包括:对于二维变化的空间变量, $P$ 形成闭合的曲线。而对于一维变化的空间变量, $P$ 为符合一定条件的离散点。以离散点表示的独立地物的分布空间永远是离散空间。

连续空间和离散空间是两个主观的概念,在实践中主要依赖于地理空间的离散性特征和连续性特征加以区分。由于地理空间中不可能存在完全均质的综合体,例在土地利用类型空间中,以“小麦地”标识的土地利用类型,在其中也许会有“房屋”用地零星分布,因此,区分连续空间和离散空间有时是很困难的<sup>[5]</sup>,并且对它们的区分往往依应用目的需要而进行。

地理空间的范化和特化、离散空间和连续空间概念的提出对建立面向数据共享的语义数据模型具有极为重要的意义,泛化和特化地理空间的概念避免了只简单地将地理空间按点、线、面进行分类的弊病,保证了完整地描述所有的地理空间体系;连续空间和离散空间则直接导致了语义数据模型基本概念——地理实体的诞生,并且对于地理实体在空间域中的函数依赖研究,从而提出GIS系统的规范化设

计理论具有重要的意义。

### 2.2 建立面向数据共享的GIS语义数据模型

我们所提出的语义数据模型建立在上述空间概念之上,核心概念是地理实体和地理特征<sup>[6,7]</sup>。

**定义** 如果  $S$  空间是关于空间变量  $a_i$  的离散空间,  $s$  为该空间的一个独立单元,则  $s$  及其在  $s$  上的空间变量  $a_i$  的值  $f(s, a_i)$  称为地理实体,地理实体是具有完整地理意义的最小的逻辑单元。地理实体具有下列特征:

(1) 地理实体是可以标识的,也就是说可以给任意地理实体唯一一标识符

(2) 地理实体是可以区分的,也就是说任意地理实体都具有一组属性,由这一组属性唯一地确定该地理实体的存在,我们称这一组属性为该地理实体的主键。

(3) 对一个地理实体的完整描述包括对其属性域及其取值、值域及其取值、地理实体基本关系和方法的描述。其中属性域是描述该地理实体所有可能的属性项的集合,值域是每一属性可能取值范围。对地理实体的属性的描述包括其空间属性及物理属性描述。概念模式是地理实体信息结构的描述,一个典型地理实体的概念模式为:

```

地理实体
{
  标识符 数据类型(例 N)
  名称 数据类型(例 S)
  几何属性 数据类型(空间数据类型)
  属性1 数据类型
  .....
  属性项 n 数据类型
  关系 数据类型
  方法
}

```

上述地理实体的定义是在离散空间中给出的,实际上可以将它们很容易地推广到连续空间中。E.Lynn Usery描述过对这一问题的处理方法<sup>[8]</sup>。

给出了地理实体的定义之后,我们可以获得一系列新的认识。针对数据共享,尤其是部门之间的数据共享,地理实体的“相等”具有特殊的意义。

**推论** 2个数据集所代表的地理实体相等,也就是2个数据集描述的是同一地理实体,必须满足的条件是2个数据集中的主键属性及其取值相等。作为判定相等的捷径,如果具有标准的地理实体标识,则标识相等意味着地理实体相等。

例如,有两概念模式  $R_1$  和  $R_2$ ,其中  $R_1$  为(空间属性,名称,标识码,人口,面积), $R_2$  为(空间属性,名称,标识码,人口)。表1和表2为对应于  $R_1$  和  $R_2$  的一组状态取值

表1 模式  $R_1$  所对应的一组状态取值

属性项	空间属性	名称	标识码	人口	面积
数据类型	GT*	Char	Integer	Integer	Float
取值	点状空间 目标	南楼	016	501	123.4

\* GT代表拓扑几何数据类型(SDTS定义)

表2 模式  $R_2$  所对应的一组状态取值

属性项	空间属性	名称	标识码	人口
数据类型	G*	Char	Integer	Float
取值	面状空间 目标	南楼	016	632

\* G代表几何数据类型(无拓扑关系)(SDTS定义)

尽管两者在属性域上的取值及各属性在各自值域上的取值均有差别,但它们由于拥有主键“名称”和“标识码”,并且其取值是相等的,因此可以认定概念模式  $R_1$  和  $R_2$  表示的是同一种地理实体,——一个名叫“南楼”、标识码为016的村庄(注意它们的空间属性的不一致性(一为点状,一为面状))。

地理特征的概念是在地理实体概念基础上给出的。

**定义** 相似的地理实体的集合及其属性的描述构成地理特征。地理特征具有内涵和外延,所有相似的地理实体集构成地理特征的外延,而描述特征属性的集合构成地理特征概念的内涵。

对地理特征的描述,与地理实体基本相似,也具有一个标识符;可以给每一地理特征赋予名称,根据应用目的不同,名称可能有多个;有一组作为主键属性的属性集合唯一确定地理特征;由于可以将地理特征看作是具有相似属性的地理实体的抽象类,因此地理实体的相似属性构成了地理特征的属性描述,其可以包括对空间属性与非空间属性的全面描述,也可以只包括对非空间属性或空间属性的描述。

适应应用目的不同,地理特征也具有相等的概念。如果存在关于地理特征的多种概念模式,其中作为主键的属性项及其取值相等,则由这一组概念模式所描述的地理特征是同一地理特征,或称地理特征的相等。地理特征相等的概念对于不同应用部门

共用同一数据集具有十分重要的意义,是部门之间数据共享的基本理论依据。

地理特征以及概括、聚合、分类、关联等数据抽

象概念<sup>[2]</sup>构成了面向数据共享的语义数据模型。

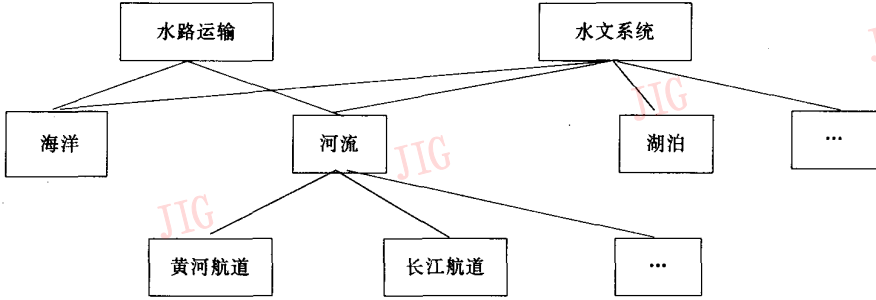


图1 借助于分类体系识别地理特征

在实践中,对地理特征的确定可以借助于分类体系,究竟将与哪一级对应的分类单位作为地理特征呢?图1中,长江作为一地理实体具有两种概念模式,一种是面向运输系统的概念模式,一种是面向水文分类的概念模式,两种模式对应于两种应用,图1中的第一级分类体现了这一应用特点,第二级分类单元“河流”,经过处理,是与应用无关的分类单位,它的下级单元便是地理实体,从图中可以看出,“河流”与其下级地理实体的关系只能为1:n关系,它与各类应用的关系也是1:n,这样我们就可以“河流”或表示同一含义的其它名词作为我们的地理特征。

给出了地理特征和地理实体的概念,我们可以建立相应的语义数据模型表达地理现实,图2为基于地理特征的地理数据共享系统示意。

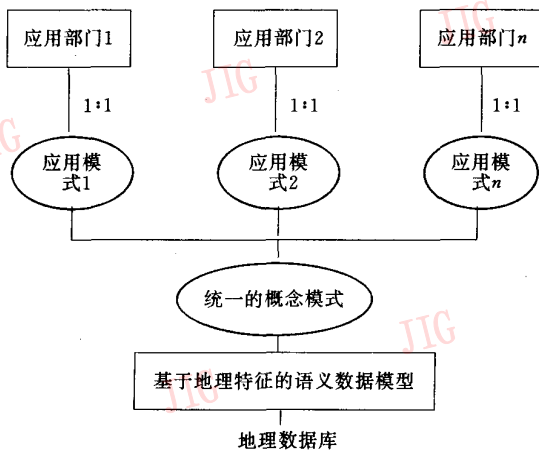


图2 GIS语义数据模型用于数据共享

### 3 应用实例研究

九五期间,我国进行国土资源、环境与地区经济信息系统(NREDIS)建设,主要工作在于将来自于林业、土地、矿产、测绘、水资源、海洋以及地区经济等部门的资源与环境数据进行集成,形成一个分布式的、逻辑上一体的NREDIS系统,其本质是要寻找各个部门之间数据相互共享的技术途径。本实例只探讨专业部门如何共享来自于测绘部门的行政区数据。

根据地理特征的定义,行政区不能再被进一步细分为同一类型的事物,因此,可以将行政区作为一类地理特征。根据数据共享中“权威数据由权威部门提供”的原则,测绘部门提供的行政区数据模式作为该地理特征的模式。如下所示:

行政区:

```

{
  Pro—Code: INTERGER, UNIQUE(省级行政区代码,唯一)
  Name: TEXT(中文省级行政区名称)
  PyName: TEXT(罗马字母拼写的省级行政区名)
  CapName: TEXT(中文省级行政区中心名)
  PyCapName: TEXT(罗马字母拼写的省级行政区中心名)
}

```

根据NREDIS总体设计之要求,大多数的专题信息都要与行政区(只要求到省级)进行集成,其中包括土地资源信息、水资源信息和森林资源信息等,为了在上述行政区特征基础上构造应用模式,分下列几步进行工作:

第一,为行政区特征添加应用属性,本项目中

采用为行政区编制应用码的方法实现。在方案中 2A2A、2B2A、2D2E 分别为行政区的 3 个应用码,代表“土地利用按行政统计区”、“水资源行政区”、和“森林资源的行政统计区”,这 3 个码使得作为地理特征的行政区变为了专题信息的统计单元。将应用码加入行政区特征模式中形成 3 个应用模式;

第二,根据具体的应用要求,统一名词术语、统一数据项要求,选取一个名称作为标准名称,在具体的应用中如果出现名词不一致的问题,采用别名处理;一个标准的行政区特征所含有的数据项并不能满足应用的需求,通过在行政区特征中增加数据项来满足应用的需求,在具体实现上可以通过定向添加关系表来实现,以利于数据的共享,图 3 为数据在逻辑上组织方法的一个例子。

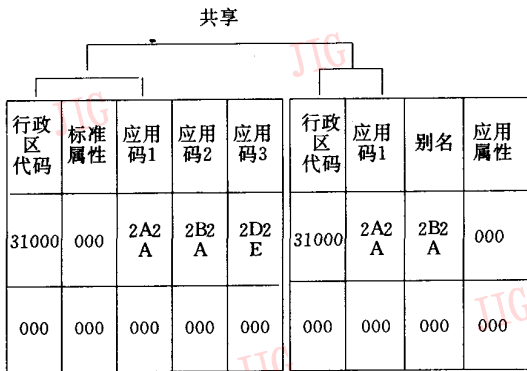


图 3 应用部门共享行政区数据

## 4 结 语

机构之间的数据共享不但需要一个合适的数据库模型作为基础,还需要对数据质量、元数据的标准化的说明。本文对满足这一需要的数据库模型进行了初步探讨,还需要做大量的工作逐步完善该模型,可以看得见的工作包括:描述自然、人文、城市地理现实的地理特征体系的建立及相关问题的研究;以该数据库为基础的共享数据库的建立及设计;GIS 数据库辅助设计工具的开发等。

## 参 考 文 献

- 1 May Yuan, Development of a global conceptual schema for inter-operable geographic information, File://opengis--paper/developing/content. htm.
- 2 陈常松. 地理信息分类体系在 GIS 语义数据模型设计中的应用. 测绘通报, 1998, (8): 17~20.
- 3 GB/T 13923-92. 国土基础信息数据分类与代码. 北京: 中国标准出版社, 1992.
- 4 Goodchild F. Geographic data modeling. Computer and Geosciences, 1992, 18(4): 401~408.
- 5 阿尔曼德 д л. 景观科学. 李世珍译. 北京: 商务书馆, 1992.
- 6 GATHA A, Tang Y, et al. A spatial data model design for feature-based geographical information system. INT. J. Geographical Information System, 1996, 10(5): 643~659.
- 7 Lynn E Usery. Category Theory and the structure of feature in Geographic information System. Cartography and Geographic information system, 1993, 20(1): 5~12.
- 8 Lynn E Usery. A feature based geographic information system model. PE&RS, 1996, 62(7): 833~828.



**陈常松** 32岁, 助理研究员, 1991年北京师范大学获硕士学位, 主要从事地理信息理论, GIS 数据共享及立法, GIS 标准化规范化等方面的研究, 1995年入中科院地理所资源与环境信息系统国家重点实验室攻读在职博士学位, 发表论文 20 多篇。



**何建邦** 中国科学院地理研究所研究员, 国际欧亚科学院院士, 博士生导师, 长期从事地理信息标准化和规范化、GIS 数据共享及立法等方面的研究工作。

## A GIS Semantic Data Model for the Purpose of Data Sharing

Chen Changsong

He Jianbang

(Chinese Academy of Surveying and Mapping, Beijing 100039) (The Institute of Geography, CAS, Beijing 100101)

**Abstract** The general way to technically implement the GIS data sharing in early days was the adoption of geographic encoding based on some classifications. The reason for this is the perception that the codes used to represent the geography is unique. But the practices in recent years have proved that the method above is of little use in data sharing. This article has given a GIS Semantic Data Model to try to solve the problem of data sharing. A theoretical frame as well as the strict definition of related concepts have been given in this paper. An application related to the design of NREDIS has been introduced to verify the applicability of the proposed model. Some further work has been listed in the last paragraph.

**Keywords** Semantic data model, Geographic entity, Geographic feature, Conceptual schema

## VTEL 发展方向 ——基于 IP 网的会议电视产品

对于渴望用更经济、更简易的网络接入方式的用户而言,基于 IP 网络上的会议电视是一项最具吸引力的技术。不久,各公司都可以利用自己的 LAN 网连接各会议电视终端,还可通过有效方式接入 ISDN 线路进行远程会议。

作为率先采用基于包传输(HDLC)技术的会议电视设备提供商,今天,根据 ITU H. 323 标准,VTEL 已成为基于 IP 网络传送数字视频通信的领导者。对于企业管理者而言,运行在 IP 网络上,基于 SNMP 协议的会议电视网管系统(Smart Video Net Manager)是用户运行维护网络的最佳方案。我们的开放式硬件平台支持以太网,令牌环网,或 ATMNIC,可轻松接入 LAN 网。

VTEL 基于 IP 的数字视频通信方案,包括会议电视系统、网关(Gateway)、多点会议服务器(MCS/IP)及其服务咨询的业务。正当其它厂商初涉这一领域时,VTEL 却以其真正完整的解决方案,满足 IP 网络环境中客户的具体需求,走在了同行业前列。

多次获得奖项的 WG500, SmartStation 和 VTEL 的全线产品 ESA 平台都支持 H. 323。为更大程度地保证您的投资,目前购买的系统(LC5000, TC2000, TC1000, WG500, SST)都可通过方便、经济地软件升级,使其具备 H. 323 功能和兼容性。此外,运行在 H. 320(ISDN)或 H. 323 模式下的系统,选择一种呼叫,就可完成 H. 320 到 H. 323 的无缝转移。VTEL 为用户提供的是一个完整方案,包括 VTEL 多点会议服务器(与 Videoserver 合作),VTEL 网关(Gateway)和网闸(Gatekeeper)。Gateway 保证了会议电视系统在 IP 网和电路交换网之间的兼容性。

为了更有效的支持我们的用户,VTEL 还将通过专业化服务组提供广泛的业务。一些具备安全意识的用户们,如需要网络设计支持或测试保证,只要接入我们提供的 LAN 环境,就可确保得到一个成功的方案。

由于视频网络在未来的几年中还将继续依附于 H. 320 技术,VTEL 还将继续延续传统的 H. 320 会议,同时,其产品更便于使用,为用户提供更加有效的通信手段。